Proc. of Int. Conf. on Emerging Trends in Engineering & Technology, IETET

Survey of Cluster based IDS using Data Mining

Manisha Kansra[1] and Pankaj Dev Chadha[2]
[1] Geeta Institute of Management and Technology/CSE, Kurukshetra ,India
Email: mishu91kansra@gmail.com
[2]Geeta Institute of Management and Technology/CSE, Kurukshetra ,India
Email:pankaj@gimtkkr.com

***Abstract*—In Information protection, intrusion detection is the act of detecting actions that attempts to compromise the confidentiality, integrity or accessibility of a resource. It plays a very important role in attack recognition, security check and network inspect. One of the most important challenges for intrusion detection are the problem of misjudgment, misdetection and lack of real time reaction to the attack. In the recent years, as the second line of defense after firewall, the intrusion detection technique has got fast development. a mixture of data mining techniques such as clustering, categorization and association rule discovery are being used for intrusion detection.**

***Index Terms*— IDS, Attack. Cluster Based IDS,Attack.**

I. INTRODUCTION

In the world of communication, security is a big concern. Most of our crucial data is stored in a computer system and in most cases we exchange it over a network. But it's not just our data transmitting over the network but different types of attacks too. These attacks can harm our stored data. Monitoring computer system and its logs (administration logs, security logs, system logs, network logs) and protecting our crucial data is necessary. For these necessities we use intrusion detection system.

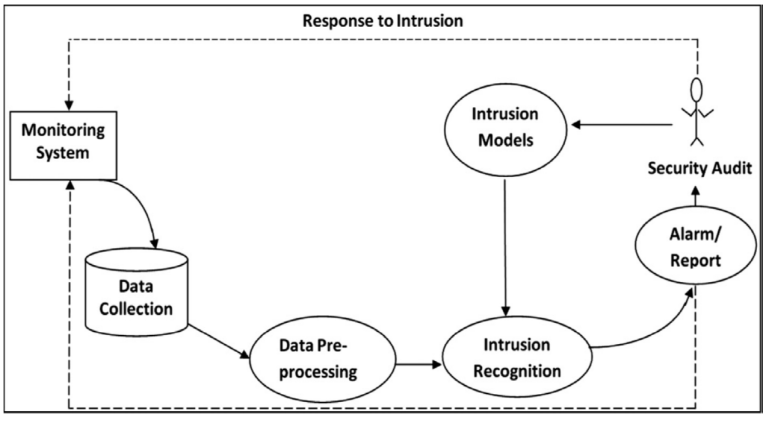

Figure 1. Flow of IDS

Grenze ID: 02.IETET.2016.5.21

An IDS essentially consists of three features: i)monitoring certain events and maintain the record of data related to that event; ii) analyzing the collected information by measuring variation in data values during a particular time period against a normal set of data; iii) generating a report about the malicious node and alert the system. Two common forms data analysis are: misuse detection and anomaly-based detection. The present day IDS are categorized based on techniques of detection as follows:

A.Mobile Agent Based IDS

One or more mobile nodes monitor the activities of all nodes and reports intrusion on its own or take a decision collaboratively. The information collected by it during collaboratively decide to vote against it or reduce it trust value. These approaches involve too much communication overhead. Life time is used to detect malicious behavior in the network. These systems help to reduce the energy consumption of sensor nodes as the load of information collection is shifted to the mobile agents.

B. Cluster Based IDS

The nodes are divided into clusters. In a cluster based approach, the cluster-head selection depends on a number of factors such as trust-worthiness, remaining power level, and average load. The nodes are monitored by the cluster head. Information about intruder nodes is passed through the gateway nodes to all the clusters.

C. Cryptography based IDS

Cryptographic techniques are used to prevent Intrusion. Intermediate nodes do not have to validate control traffic. Different results obtained by route discovery method are used as alternative routes. False route and topological information can be detected.

D. Neigh hood Watch IDS

A node checks the number of packets received by a forwarding the actual number of packets forwarded by it. Any deviation from normal trend leads to a suspect tag for the node. Watch dog mechanism based on promiscuous mode operation of the ad hoc nodes, has been the fundamental assumption in any trust computational model.

E. Cross-feature Analysis IDS

These IDS solutions utilize Cross-Layer co-operation. Behaviour of nodes is classified in some classes depending on some pre decided features. A node is decided to be malicious or not depending on its features. In this method, primary misbehaviour detection system (MDS-p) in MAC layer creates a trust list depending on some decided trust metrics and observing misbehaviour. Trust level is computed for each node and a node is identified as intruder if the trust level falls below a threshold.

F. Collaborative IDS

The decision of a node to be termed an intruder is taken collaboratively andgenerally in promiscuous mode. These approaches involve too much communication overhead.

II. Intrusion Detection System based on Data Mining

Using data mining into intrusion detection structure improve the performance has become one of the major concerns in the research of intrusion detection. Data mining usually refers to
the process of removing descriptive models from large stores of data. The current rapid development in data mining has made available a change of algorithm, drawn from the field of information, pattern recognition, machine learning. Some types of algorithms are particularly useful for mining data. Intrusion detection has great practical drive and application value therefore intrusion detection based on data mining cannot stop at the theoretical investigation. The joint use of several data mining methods can effectively improve data processing speed and quality. Data mining provides decision support for intrusion management. It also discovered unidentified patterns of attack or intrusions. In this way it helps intrusion detection system for detecting new vulnerabilities and intrusions. Data mining can improve variant detection rate, control false alarm rate and reduce false dismissals. There are many data mining methods. Data mining can be divided into four types: association analysis, series analysis, classification analysis and cluster analysis.

Association provides simple but valuable description form for the rule mode in data-mining i.e. describe invasion of behaviour patterns. Classification maps the data item into one of the several predefined classes.

III. Cluster Analysis Methods in Data Mining

Cluster analysis is a very important data mining technology to divide the data object into several meaningful subclasses, so that the members from the same clusters are quite similar and members from different clusters are quite different from each other [7]. Therefore, this method is applied for classifying log data and detecting intrusions. Clustering is an unsupervised learning system of data mining that takes unlabeled data points and tries to group them according to their similarity. In unsupervised approach there is no need of previous knowledge about training data whereas in supervised approach, given a set of normal data need to train in order to detect whether the test data belongs to normal or anomalous behaviour [8].The general steps for clustering feature extraction from sample data where input is sample data and output is matrix. Then implementation of clustering algorithm to access cluster genealogy diagram i.e. to reflect all the classification. After obtaining a cluster genealogy diagram, farther first domain experts will is applied to the anomaly detection problem the based clustering with low time complexity and fast convergence which is very important in intrusion detection due to large size of network traffic audit dataset [8]. K-means algorithm divides N data object into K clusters. Theo objects in the same clustering have higher similarity while objects in different clustering have smaller similarity. It is a dynamic clustering based on standard measure function. K-means algorithm divides N vectors into K classes. Usually start with an initial partition then use an iterative control strategy to optimize an objective function [8]. K-means represents a type of use full clustering techniques by competitive learning which is also one of the promising techniques in intrusion detection.

IV. Related Work

*Kalpana Jaswal et al*An intrusion detection classification is an application that provides protection from malicious activities or policy violations and generates various rules to defend computer security. Intrusion detection system can be designed and developed on any platform but for its better functionality we are using data mining technique. In past years, many techniques have been introduced to manage the detection rate. Earlier, in the initial stages of its designing, hardware had to be installed to detect and monitor the system. But, with the help of data mining it has become easier to work with software and algorithm development. In the recent trends, many new algorithms have been introduced to increase its efficiency. They are categorized under machine learning algorithms: supervised, unsupervised and hybrid. Though hybrid has not yet been categorized finely but various authors have used it by merging different machine learning algorithms

*S.V.Shirbhate et al*The study, analysis and exploration of recent expansion of data mining applications such as categorization and clustering is one of the needs for machine learning algorithms to be applied to great scale data will guide to acquire the direction of future research. It would be prospect demand in IDS for detecting the intrusions in mobile network. presents the comparison of different clustering techniques. Also focus on the effect of Principal Constituent Analysis filter on these clustered based methods.

*Chakchai Soet al*Due to a rapid growth of Internet, the number of network attacks has risen important to the essentials of network intrusion detection systems (IDS) to safe and sound the network. With heterogeneous accesses and enormous traffic volumes, more than a few pattern identification techniques have been brought into the research area. Data Mining is one of the analyses which many IDSs have adopt as an attack recognition scheme the classification methodology including attribute and data selection was tired based on the well–known classification schemes, i.e., Decision Tree, Ripper Rule,Naïve Bayes, *k*–Nearest–Neighbour, and Support Vector Machine, for intrusion detection analysis using both KDD CUP dataset and recent HTTP BOTNET attacks. presentation of the evaluation was measured using recent Weka tools with a standard cross–validation and confusion matrix.

*Mouaad KEZIH*Intrusions detections systems from point of view of security policy are a second line of defense; they have a supervisory role to observe the activities of our network or hosts to identify attacks in actual time. In our days, electronics attacks can cause a very destructive damage for nations which make necessary the use of completed security policy to minimize the probable threats. IDS it is a very important element to resist against this vulnerability, (KDD) CUP 99 and a Data Mining Tools Waikato Environment for Knowledge Analysis (WEKA) to combine the advantages of an intrusion detection algorithm (PART) and

two techniques of Dimensionality Reduction (best first search and genetic search), to estimate our works, we applied the proposed combined technique, and we check the results by using a several evaluations parameters.

V. ATTACKS

A. Selfish Behaviors

Selfish node does not cooperate in any network functions and exploits the services of the network for its advantage, in order to save its own resources such as battery life. While such misbehaviour may not be launched with explicitly bad intentions, it can lead to serious disruptions in network communications such as tall route discovery delays and dropped data packets. To render its selfish behaviour undetectable, selfish node can use multiple identities (e.g. Sybil attack) [8].

1) Routing messages dropping

Selfish node intentionally drops routing packets that are not destined for it, or forwards those packets but with a time-to live(TTL) of 0 to prevent the creation of routes. There by the selfish node can avoid forwarding many subsequent data packets [8]. This attack may reduce the network performance and prevent end-to-end communications between nodes, if the dropping node is at a critical point in the network [2].

2) Routing metric inflation

Selfish node may make routes through itself appear longer than they actually are, by increasing hop counts so the source nodes are more probably to select other routes that
seem to be shorter. Selfish node can also generate packets that advertise arbitrarily high distance to a given destination, or stops announcing updates that contain better routes that
pass through it. In routing protocols that suppress duplicate packets, such as DSR [6] and AODV [1], where each node forwards only the first received packet and deletes any later copies of the same packet. Selfish node can break this rule by waiting to receive several duplicates and then forwards the packet with the highest routing metric, decreasing the probability of being selected in the discovered route [2].

B. Malicious attacks

1) Modification

Modification is the most common attack; in which malicious node modifies the content fields of routing packets that transit through it. A malicious node could modify packets before rebroadcasting them, so that they include less attractive metrics, false addresses, and fake hop
count in order to redirect network traffic. This attack can cause severe routing disruptions such as; conflicted and suboptimal routes, erroneous routing table, network partition and lose of connectivity [7].

2) Fabrication

This attack refers to the generation of faked routing messages, in order to disrupt network operation or to deplete other nodes' resources. Such attack is difficult to detect [7].

3) Impersonation

Also called spoofing attack, it usually constitutes the first step in the majority of attacks. The malicious node hides its real identity and takes legitimate node's identity, thus it canreceive all the messages destined to this node and gain access to the network. This attack can also be used for creating loops in order to isolate a target node from the rest of the network [2].

4) Black hole

This attack exploits the vulnerabilities of routing protocols and it is carried out in two steps. First, the malicious node attracts traffic through itself by advertising better routes to the requested destinations. Afterward, the malicious node drops all the data or control packets passing through it without any forwarding [2].

5) Gray hole

This attack is a refined form of black hole attack, in which a malicious node drops only chosen packets and forwards the others, depends on the source or the destination ofpackets. Another kind of gray hole may behave maliciously for a given period by dropping all packets then switch to normal behaviour later. This attack defeats trust-based mechanisms and makes the detection of malicious node more difficult to achieve [1].

6) Wormhole

Also called tunneling attack, it is one of the most sophisticated attacks in MANETs. In this attack, a malicious node capture packet from one location in a network and tunnels them through an out-of-band channel to another malicious node located several hops gone, which replays them to its neigh boring nodes,. tunnel among the malicious nodes is actually faster than links between legitimate nodes, so the tunneled packets arrive sooner than packets through other routes. Therefore, the malicious nodes are more likely to be included in the route and take an advantage for future attack. Detection of wormhole attack is generally difficult, and requires the use of an unalterable and independent physical metric, such as time delay or geographical location [8].

VI. Taxonomy of Intrusion Detection Systems

Intrusion Detection System can be classified in three ways. This taxonomy of Intrusion Detection System is given as follows:

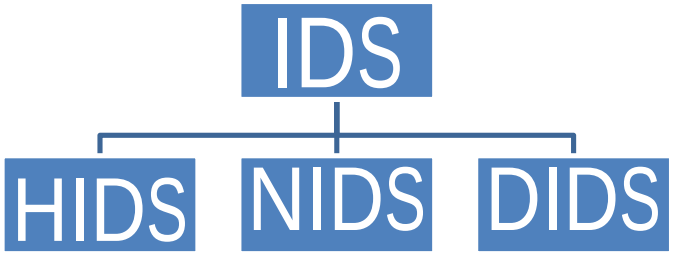

Figure 2. Taxonomy of IDS

A. Host Based Intrusion Detection System (HIDS)

Host based Intrusion Detection Systems have sensors which focuses only on single host for the detection of the intrusion. A HIDS monitors the incoming and outgoing packets from the host and alerts the user or administrator of suspicious activity if detected any.

Advantages

1. Host based IDS monitors an individual system in the network. Thus, it can detect potentially dangerous activities on a single system in a network.
2. As, HIDS runs on a single machine it remains isolated from the network traffic.
3. Comparatively more acute.
4. There agents can collect more information.

Disadvantages

1. There are some HIDS which can detect only for certain type of systems.
2. The generation of alerts is slow.
3. Resource usage by agents slows down the speed of operations.
4. Ineffective during DoS attacks.

B. Network Based Intrusion Detection System (NIDS)

Network based IDS have sensors which detect the intrusions over the network. NIDS are placed at a strategic point or points within the network to observe incoming and outgoing traffic all devices on the network.

Advantages

1. It can have an eye over whole or part of network.
2. It also monitors general traffic problems and provides troubleshooting.
3. Effective during DoS attacks.

Disadvantages

1. One of the limitation of NIDS is high frequency false positives.
2. The TCP/IP packets visiting the IDS come unordered. Sometimes the packet arriving at first appears like an ordinary packet and gets undetected by the IDS and thus can crash the IDS.
3. The performance gets affected by the network traffic.

C. Distributed Intrusion Detection System (DIDS)

Distributed IDS integrates both types of sensors. A DIDS consists of several IDS over a great network all of which communicate with each other, or with a central server that facilitates advanced network monitoring [4].

Advantages

1. It has features of both HIDS and NIDS.
2. In this kind of IDs multiple IDS are deployed over the network.
3. It reduces response time.

Disadvantages

1. It is difficult to implement as it has complex structure
2. It is expensive.

VII. CONCLUSION

In this paper we provided a detailed study of IDS that occurred in Data Mining. Thus there will likely be obstacles in developing an efficient solution Intrusion detection system is an area of active research for above fifteen years. Current commercial intrusion detection systems make use of misuse detection. As such, they completely are short of the ability to detect new attacks. It is impossible to prevent security violation completely by using the exciting security technologies. Accordingly, Intrusion Detection is an important component of network security.

REFERENCES

[1] Mouaad KEZIH Mahmoud TAIBI, "Evaluation Effectiveness of Intrusion Detection System with Reduced Dimension Using Data Mining Classification Tools," 2nd International Conference on Systems and Computer Science (ICSCS) Villeneuve d'Ascq, France, August 26-27, 2013.

[2] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu and Ali A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set,"Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA), 2009.

[3] M.Govindaraja,Rlvl.Chandrasekaran, "Intrusion Detection Using k-Nearest Neighbor,"vol. 2, Issue 3, IEEE, 2009

[4] Tayeb Kenaza, Abdelhalim Zaidi, "Clustering approach for false alerts reducing in behavioral based intrusion detection systems," vol. 4, Issue 5, IEEE, 2010.

[5] Chang-Tien Lu, Arnold P. Boedihardjo, Prajwal Manalwar, "Exploiting Efficient Data Mining Techniques to Enhance Intrusion Detection Systems," vol 2, Issue 3, IEEE, 2005

[6] Meng Jianliang Shang Haikun Bian Ling, "The Application on Intrusion Detection Based on K-means Cluster Algorithm,"IEEE International Forum on Information Technology and Applications,2009.

[7] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification," 7th International Conference on IT in Asia (CITA),2011.

[8] Slobodan Petrovic, Gonzalo A´ lvarez2, Agust´ın Orfila3, and Javier Carbo, "Labelling Clusters in an Intrusion Detection System Using a Combination of Clustering EvaluationTechniques,"Proceedings of the 39th Hawaii International Conference on System Sciences,2006.

[9] Cuixiao Zhang; Guobing Zhang; Shanshan Sun, "A Mixed Unsupervised Clustering-based Intrusion Detection Model,"Third International Conference on Genetic and Evolutionary Computing,2009.

[10] Tingting Wang, ZhaobinLiu, Yi Chen, Yujie Xu, "Load Balancing Task Scheduling based on Genetic Algorithm in Cloud Computing" IEEE 12th International Conference on Dependable, Autonomic and Secure Computing,2014.

[11] Kalpana Jaswal, Praveen Kumar, "Design and Development of a Prototype Application for Intrusion Detection using Data mining,"IEEE ,2015.

[12] S.V.Shirbhate, Dr.S.S.Sherekar, "Performance Evaluation of PCA Filter In Clustered Based Intrusion Detection,"International Conference on Electronic Systems, Signal Processing and Computing Technologies IEEE,2014.

[13] Chakchai So, "An Evaluation of Data Mining Classification Models for Network Intrusion Detection," vol.2, Issue 3, IEEE, 2014.